Discriminating quantum states: the multiple Chernoff distance

Ke Li

IBM T.J. Watson Research Center
and Massachusetts Institute of Technology

Abstract

We consider the problem of testing multiple quantum hypotheses $\{\rho_1^{\otimes n}, \ldots, \rho_r^{\otimes n}\}$, where an arbitrary prior distribution is given and each of the r hypotheses is n copies of a quantum state. It is known that the minimal average error probability P_e decays exponentially to zero, that is, $P_e = \exp\{-\xi n + o(n)\}$. However, this error exponent ξ is generally unknown, except for the case that $r = 2$.

In this paper, we solve the long-standing open problem of identifying the above error exponent, by proving Nussbaum and Szkoła's conjecture that $\xi = \min_{i \neq j} C(\rho_i, \rho_j)$. The right-hand side of this equality is called the multiple quantum Chernoff distance, and $C(\rho_i, \rho_j) := \max_{0 \leq s \leq 1}\{-\log \operatorname{Tr} \rho_i^s \rho_j^{1-s}\}$ has been previously identified as the optimal error exponent for testing two hypotheses, $\rho_i^{\otimes n}$ versus $\rho_j^{\otimes n}$.

The main ingredient of our proof is a new upper bound for the average error probability, for testing an ensemble of finite-dimensional, but otherwise general, quantum states. This upper bound, up to a states-dependent factor, matches the multiple-state generalization of Nussbaum and Szkoła's lower bound. Specialized to the case $r = 2$, we give an alternative proof to the achievability of the binary-hypothesis Chernoff distance, which was originally proved by Audenaert et al.

Email: carl.ke.lee@gmail.com

Supported by NSF grants CCF-1110941 and CCF-1111382.

AMS 2000 subject classifications. 62P35, 62G10.

Key words and phrases. quantum state discrimination, quantum hypothesis testing, error exponent, quantum Chernoff distance, multiple hypotheses

1 Introduction

A basic problem in information theory and statistics, is to test a system that may be prepared in one of r random states. Treated in the framework of quantum mechanics, the testing is performed via quantum measurement, and the physical states are described by density matrices $\omega_1, \omega_2, \ldots, \omega_r$, namely, positive semidefinite Hermitian matrices of trace 1. It is a notable fact that, when ω_i's commute, the problem reduces to classical statistical testing, among r probability distributions that are given by the arrays of eigenvalues of each of the density matrices. However, the generally noncommutative feature makes quantum statistics much richer than its classical counterpart.

Our main focus in the current paper will be on the asymptotic setting. Let the tensor product state $\rho^{\otimes n}$ denotes n independent copies of ρ, in analogy to the probability distribution of i.i.d. random variables. We are interested in the asymptotic behavior of the average error P_e, in discriminating a set of quantum states $\{\rho_1^{\otimes n}, \ldots, \rho_r^{\otimes n}\}$, when an arbitrary prior that is independent of n is given. Parthasarathy showed that P_e decays exponentially, that is, $P_e = \exp\{-\xi n + o(n)\}$ [34]. However, to date the optimal error exponent ξ, as a functional of the states $\rho_1, \ldots, \rho_r$, is generally unknown.

Significant achievements have been made for the case of testing two quantum hypotheses ($r = 2$). In two breakthrough papers, [1] and [29], it has been established that the optimal error exponent in discriminating $\rho_1^{\otimes n}$ and $\rho_2^{\otimes n}$, equals the quantum Chernoff distance

$$C(\rho_1, \rho_2) := \max_{0 \le s \le 1} \{-\log \operatorname{Tr} \rho_1^s \rho_2^{1-s}\}.$$

Audenaert et al in [1] solved the achievability part, in the meantime Nussbaum and Szkoła in [29] proved the optimality part. This provides the quantum generalization of the Chernoff information as the optimal error exponent in classical hypotheses testing [9]; see also [10].

The solution for the general cases $r > 2$ is still lacking and it does not follow from the binary case directly. The optimal tests, as analogs of the classical maximum likelihood decision rule, have been formulated in the 1970s. For discriminating two states it has an explicit expression known as the Holevo?Helstrom test [16, 21], and indeed, the proof in [1] relies on a nontrivial application of this Holevo?Helstrom test. In contrast, for discriminating multiple quantum states the corresponding optimal measurement can only be formulated in a very complicated, implicit way [20, 41]. Such a situation illustrates the difficulty in dealing with the asymptotic error exponent,

for the multiple case $r > 2$. Intuitively, competitions among pairs make the problem complicated.

Nussbaum and Szkoła introduced the multiple quantum Chernoff distance

$$C(\rho_1, \dots, \rho_r) := \min_{(i,j): i \neq j} C(\rho_i, \rho_j),$$

and conjectured that it is the optimal asymptotic error exponent, in discriminating quantum states $\rho_1^{\otimes n}, \dots, \rho_r^{\otimes n}$ [30, 31, 32]. This is in full analogy to the existing results in classical statistical hypothesis testing [23, 36, 37, 39]. Significant progress has been made towards proving this conjecture. Besides the case of commuting states which reduces to the classical situation, it has been proven to be true in several interesting special cases. These include when the supporting spaces of the states $\rho_1, \dots, \rho_r$ are disjoint [32], and when one pair of the states is substantially closer than the other pairs, in Chernoff distance [28, 2]. In general, Nussbaum and Szkoła showed that the optimal error exponent ξ in testing multiple quantum hypotheses, satisfies $C/3 \leq \xi \leq C$ [32], and Audenaert and Mosonyi recently strengthened this bound, showing that $C/2 \leq \xi \leq C$ [2].

In this paper, we shall prove the aforementioned conjecture, that is, we show that the long-sought error exponent in asymptotic quantum (multiple) state discrimination, is given by the (multiple) quantum Chernoff distance. Besides, as a main ingredient of the proof we derive a new upper bound for the optimal average error probability, for discriminating a set of finite-dimensional, but otherwise general, quantum states. This one-shot upper bound has the advantage that, up to a states-dependent factor, it coincides with a multiple-state generalization of Nussbaum and Szkoła's lower bound [29].

Before concluding this section, we review the relevant literature. Asymptotics of statistical hypothesis testing is an important topic in statistics and information theory, and is especially useful in identifying basic information quantities. We refer the interested readers for a partial list of classical results to [6, 9, 10, 13, 14, 19, 23], and quantum results to [1, 3, 5, 8, 18, 24, 27, 29, 33, 38]. The optimal or approximately optimal average error in quantum state discrimination, and the corresponding tests to achieve it, is a basic problem in quantum information theory and has attracted extensive study; see, for example, [2, 4, 15, 16, 20, 21, 22, 35, 40, 41].

The remainder of this paper is organized as follows. After introducing some basic notations, concepts and the relevant aspects of the quantum formalism in Section 2, we present the main results in Section 3. Section 4 is dedicated to the proofs. At last, in Section 5, we conclude the paper with

some discussion and open questions.

2 Notation and preliminaries

Let $\mathcal{B}(\mathcal{H})$ denote the set of linear operators on a complex, finite-dimensional Hilbert space $\mathcal{H}$. Let $\mathcal{P}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ be the set of positive semidefinite matrices, and $\mathcal{D}(\mathcal{H}) := \{\omega : \omega \in \mathcal{P}(\mathcal{H}), \text{and} \operatorname{Tr}\omega = 1\}$ is the set of density matrices. We say a matrix $A \geq 0$ if $A \in \mathcal{P}(\mathcal{H})$, and $A \geq B$ if $A - B \geq 0$. The dimension of the Hilbert space $\mathcal{H}$ is denoted as $|\mathcal{H}|$. $\mathbb{1}$ denotes the identity matrix. We use the Dirac notation $|v\rangle \in \mathcal{H}$ to denote a unit vector, $\langle v|$ its conjugate transpose, and $\langle v|w\rangle$ the inner product. A Hermitian matrix X can be written in the spectral decomposition form: $X = \sum_i \lambda_i Q_i$, where λ_i's satisfying $\lambda_i \neq \lambda_j$ for $i \neq j$ are the eigenvalues, and Q_i's satisfying $Q_i Q_j = \delta_{ij} Q_i$ and $\sum_i Q_i = \mathbb{1}$ are the orthogonal projectors onto the eigenspaces. $\operatorname{supp}(X)$ is the supporting space of X and is spanned by all the eigenvectors with non-zero eigenvalues, $\{X > 0\} := \sum_{i:\lambda_i > 0} Q_i$ represents the projector onto the positive supporting space of X, and $\Omega(X) := |\{\lambda_i\}_i|$ denotes the number of eigenspaces, or distinct eigenvalues. For a subspace $S \subset \mathcal{H}$, $\operatorname{proj}(S)$ is the projector onto S. The sum of two subspaces $S_1, S_2 \subset \mathcal{H}$, is defined as $S_1 + S_2 := \{u + v | u \in S_1, v \in S_2\}$. When we say the overlap between two subspaces S_1 and S_2, we mean the maximal overlap between two unit vectors from each of them: $\max\{|\langle v|w\rangle| : |v\rangle \in S_1, |w\rangle \in S_2\}$.

We briefly review some aspects of the quantum formalism, relevant in this paper. Every physical system is associated with a complex Hilbert space, which is called the state space. The states of a system are described by density matrices. Pure states are of particular interest, and are represented by rank-one projectors, or simply the corresponding unit vectors. Throughout this paper, we are concerned with quantum states of a finite system, associated with a finite-dimensional Hilbert space. A density matrix ω can be decomposed as the sum of an ensemble of pure states, that is, $\omega = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, with $\{p_i\}$ a probability distribution. An intuitive understanding is that pure states represent "deterministic events", and a density matrix is the quantum analogue of a probability distribution over these events. However, note that this decomposition is not unique, and non-orthogonal pure states are not perfectly distinguishable.

The procedure of detecting the state of a quantum system is called quantum measurement, which, in the most general form, is formulated as positive operator-valued measure (POVM), that is, $\mathcal{M} = \{M_i\}_i$, with the POVM elements satisfying $0 \leq M_i \leq \mathbb{1}$ and $\sum_i M_i = \mathbb{1}$. When performing the

measurement on a system in the state ω, we get outcome i with probability $\mathrm{Tr}(\omega M_i)$. Projective measurements, or von Neumann measurements, are special situations of POVMs, where all the POVM elements are orthogonal projectors: $M_i M_j = \delta_{ij} M_i$, with δ_{ij} the Kronecker delta function.

Suppose a physical system, also called an information source, is in one of a finite set of hypothesized states $\{\omega_1, \ldots, \omega_r\}$, with a given prior $\{p_1, \ldots, p_r\}$. For convenience, we denote them as a normalized ensemble $\{A_1 := p_1\omega_1, \ldots, A_r := p_r\omega_r\}$. To determine the true state, we make a POVM measurement $\{M_1, \ldots, M_r\}$, and infer that it is in the state ω_i if we get outcome i. The average (Bayesian) error probability is

$$P_e\left(\{A_1, \ldots, A_r\}; \{M_1, \ldots, M_r\}\right) := \sum_{i=1}^{r} \mathrm{Tr}\, A_i(\mathbb{1} - M_i). \tag{1}$$

Minimized over all possible measurements, this gives the optimal error probability

$$P_e^*\left(\{A_1, \ldots, A_r\}\right) := \min\left\{\sum_{i=1}^{r} \mathrm{Tr}\, A_i(\mathbb{1} - M_i) : \mathrm{POVM}\, \{M_1, \ldots, M_r\}\right\}. \tag{2}$$

We note here that the definitions (1) and (2) apply, as well, to a non-normalized ensemble of quantum states $\{A_1, \ldots, A_r\}$ which only satisfies the constraint $(\forall i)\, A_i \geq 0$. In this case, P_e and P_e^* may not have a clear meaning but sometimes can be useful.

In the asymptotic setting where ω_i is replaced by the tensor product state $\rho_i^{\otimes n}$, we are interested in the behavior of the optimal error P_e^*, as $n \to \infty$. An important quantity characterizing this asymptotic behavior, is the rate of exponential decay, or simply *error exponent*

$$\liminf_{n\to\infty} \frac{-1}{n} \log P_e^*\left(\{p_1\rho_1^{\otimes n}, \ldots, p_r\rho_r^{\otimes n}\}\right).$$

3 Results

Our main result is the following Theorem 1. Recall that, for the case $r = 2$ of testing two hypotheses, it has been proven nearly a decade ago in 2006; see [1] and [29].

Theorem 1. *Let $\{\rho_1, \ldots, \rho_r\}$ be a finite set of quantum states on a finite-dimensional Hilbert space $\mathcal{H}$. Then the asymptotic error exponent for testing*

$\{\rho_1^{\otimes n}, \dots, \rho_r^{\otimes n}\}$, *for an arbitrary prior* $\{p_1, \dots, p_r\}$, *is given by the multiple quantum Chernoff distance:*

$$\lim_{n\to\infty} \frac{-1}{n} \log P_e^* \left(\{p_1\rho_1^{\otimes n}, \dots, p_r\rho_r^{\otimes n}\}\right) = \min_{(i,j):i\neq j} \max_{0\leq s\leq 1} \left\{-\log \operatorname{Tr} \rho_i^s \rho_j^{1-s}\right\}. \quad (3)$$

The optimality part, that is, the left-hand side of equation (3) being upper bounded by the right-hand side, follows easily from the optimality of the binary case $r = 2$ [29]; see the argument in [31]. Roughly speaking, this is because, discriminating an arbitrary pair within a set of quantum states is easier than discriminating all of them. On the other hand, the achievability part is the main difficulty in proving Theorem 1. In [1], Audenaert et al employed the Holevo?Helstrom tests $\left(\{\rho_1^{\otimes n} - \rho_2^{\otimes n} > 0\}, \mathbb{1} - \{\rho_1^{\otimes n} - \rho_2^{\otimes n} > 0\}\right)$ to achieve the binary Chernoff distance in testing $\rho_1^{\otimes n}$ versus $\rho_2^{\otimes n}$. However, to date we do not have a way to generalize the method of Audenaert et al to deal with the $r > 2$ cases, even though there is the multiple generalization of the Holevo?Helstrom tests [20, 41]; see discussions in [32] and [2] on this issue. Here, using a conceptually different method, we derive a new upper bound for the optimal error probability of equation (2). This one-shot error bound, as stated in Theorem 2, works for testing any finite number of finite-dimensional quantum states, and when applied in the asymptotics for i.i.d. states, accomplishes the achievability part of Theorem 1.

Our method is inspired by the previous work of Nussbaum and Szkoła [32]. It is shown in [32] that if the supporting spaces of the hypothetic states $\rho_1, \dots, \rho_r$ are pairwise disjoint (this means that the supporting spaces of $\rho_1^{\otimes n}, \dots, \rho_r^{\otimes n}$ are asymptotically highly orthogonal), then the Gram-Schmidt orthonormalization can be employed to construct a good measurement, which achieves the error exponent of Theorem 1. Here to prove Theorem 1 for general hypothetic states, we find a way to remove a subspace from each eigenspace of the states $\rho_1^{\otimes n}, \dots, \rho_r^{\otimes n}$. Then we show that, on the one hand this removal will cause an error that matches the right-hand side of equation (3) in the exponent, and on the other hand the pairwise overlaps between the supporting spaces of $\rho_1^{\otimes n}, \dots, \rho_r^{\otimes n}$ are made sufficiently small, such that the Gram-Schmidt orthonormalization method is applicable. For the sake of generality, we will actually realize these ideas for general nonnegative matrices $A_1, \dots, A_r$, yielding the following Theorem 2.

Theorem 2. *Let* $A_1, \dots, A_r \in \mathcal{P}(\mathcal{H})$ *be nonnegative matrices on a finite-dimensional Hilbert space* $\mathcal{H}$*. For all* $1 \leq i \leq r$*, let* $A_i = \sum_{k=1}^{T_i} \lambda_{ik} Q_{ik}$ *be the spectral decomposition of* A_i*, and write* $T := \max\{T_1, \dots, T_r\}$*. There exists*

a function $f(r,T)$ such that

$$P_e^*(\{A_1,\ldots,A_r\}) \leq f(r,T) \sum_{(i,j):i<j} \sum_{k,\ell} \min\{\lambda_{ik},\lambda_{j\ell}\} \operatorname{Tr} Q_{ik}Q_{j\ell} \tag{4}$$

and we have $f(r,T) < 10(r-1)^2T^2$.

Our upper bound of equation (4), up to an r- and T-dependent factor, coincides with the multiple-state generalization of the lower bound of Nussbaum and Szkoła [29]. To see this, using the result in [35], we easily generalize the bound obtained in [29] and get

$$P_e^*(\{A_1,\ldots,A_r\}) \geq \frac{1}{2(r-1)} \sum_{(i,j):i<j} \sum_{k,\ell} \min\{\lambda_{ik},\lambda_{j\ell}\} \operatorname{Tr} Q_{ik}Q_{j\ell}. \tag{5}$$

In the case that $r = 2$, it is interesting to compare equation (4) with the upper bound of Audenaert et al [1]:

$$P_e^*(\{A_1,A_2\}) \leq \min_{0\leq s\leq 1} \operatorname{Tr} A_1^s A_2^{1-s}. \tag{6}$$

While we see that our bound is stronger, in the sense that

$$\sum_{k,\ell} \min\{\lambda_{1k},\lambda_{2\ell}\} \operatorname{Tr} Q_{1k}Q_{2\ell} \leq \min_{0\leq s\leq 1} \operatorname{Tr} A_1^s A_2^{1-s} \tag{7}$$

is always true, we also notice that it is weaker because it has an additional multiplier depending on the number of eigenspaces of the two states.

4 Proofs

This section is dedicated to the proofs of Theorem 1 and Theorem 2. At first, we present a definition and some necessary lemmas in Section 4.1. Then we construct the measurement for discriminating multiple quantum states in Section 4.2. Using this measurement, we prove Theorem 2 in section 4.3. At last, built on Theorem 2, Theorem 1 will be proven in Section 4.4.

4.1 Technical preliminaries

We begin with the definition of the operation "ϵ-subtraction" between two projectors or two subspaces. This operation, say, for two subspaces S_1 and S_2, reduces the overlap between them by removing a subspace from S_1, actually, in the most efficient way. It will constitute a key step in the construction of measurement.

Definition 3 (ϵ-subtraction)**.** *Let S_1, S_2 be two subspaces of a Hilbert space $\mathcal{H}$. Let $P_1, P_2 \in \mathcal{P}(\mathcal{H})$ be the projectors onto S_1 and S_2, respectively. Write $P_1P_2P_1$ in the spectral decomposition, $P_1P_2P_1 = \sum_x \lambda_x Q_x$, with Q_x being orthogonal projectors and $\sum_x Q_x = \mathbb{1}_{\mathcal{H}}$. For $0 \leq \epsilon \leq 1$, the ϵ-subtraction of P_2 from P_1 is defined as*

$$P_1 \ominus_\epsilon P_2 := P_1 - \sum_{x:\lambda_x \geq \epsilon^2, \lambda_x \neq 0} Q_x. \tag{8}$$

Accordingly, the ϵ-subtraction between subspaces is defined as

$$S_1 \ominus_\epsilon S_2 := \mathrm{supp}(P_1 \ominus_\epsilon P_2). \tag{9}$$

Note that in equation (8) the constraint $\lambda_x \neq 0$ makes sense only when $\epsilon = 0$. The following lemma states some basic properties of the ϵ-subtraction.

Lemma 4. *Let S_1, S_2 be two subspaces of a Hilbert space $\mathcal{H}$. Let $P_1, P_2 \in \mathcal{P}(\mathcal{H})$ be the projectors onto S_1 and S_2, respectively. Write $S_1' = S_1 \ominus_\epsilon S_2$, and $P_1' = P_1 \ominus_\epsilon P_2$. Then*

1. *S_1' is a subspace of S_1; P_1' is a projector, and $0 \leq P_1' \leq P_1$.*
2. *S_1' has bounded overlap with S_2:*
$$\max_{|v_1\rangle \in S_1', |v_2\rangle \in S_2} |\langle v_1|v_2\rangle| \leq \epsilon,$$
where the maximization is over unit vectors $\langle v_1|v_1\rangle = \langle v_2|v_2\rangle = 1$.
3. *For $0 < \epsilon \leq 1$, we have*
$$\mathrm{Tr}(P_1 - P_1') \leq \frac{1}{\epsilon^2} \mathrm{Tr}\, P_1 P_2.$$

Proof. Let $P_1P_2P_1 = \sum_x \lambda_x Q_x$ be the spectral decomposition of $P_1P_2P_1$, with $0 \leq \lambda_x \leq 1$.

1. Obviously, $\mathrm{supp}(P_1P_2P_1) \subseteq S_1$. Thus the following three projectors satisfy

$$\sum_{x:\lambda_x \geq \epsilon^2, \lambda_x \neq 0} Q_x \leq \sum_{x:\lambda_x \neq 0} Q_x \leq P_1. \tag{10}$$

This, together with Definition 3, implies that P_1' is a projector and satisfies $0 \leq P_1' \leq P_1$. The fact that S_1' is a subspace of S_1, follows directly.

2. It follows from equation (10) and Definition 3 that we can write S_1' as

$$S_1' = \left(\bigoplus_{x:0<\lambda_x<\epsilon^2} \mathrm{supp}(Q_x) \right) \bigoplus \left(\mathrm{supp}(Q_x)|_{\lambda_x=0} \bigcap S_1 \right).$$

That is to say, S_1' is the direct sum of the eigenspaces of $P_1P_2P_1$ with corresponding eigenvalues in the interval $(0, \epsilon^2)$, together with a subspace of the kernel of $P_1P_2P_1$. So, for any unit vectors $|v_1\rangle \in S_1'$, $|v_2\rangle \in S_2$,

$$|\langle v_1|v_2\rangle| = \sqrt{\langle v_1|v_2\rangle\langle v_2|v_1\rangle} \leq \sqrt{\langle v_1|P_2|v_1\rangle} = \sqrt{\langle v_1|P_1P_2P_1|v_1\rangle} \leq \epsilon.$$

3. This inequality can be verified as follows.

$$\begin{aligned}
\mathrm{Tr}(P_1 - P_1') &= \mathrm{Tr} \sum_{x:\lambda_x \geq \epsilon^2} Q_x \\
&\leq \mathrm{Tr} \sum_{x:\lambda_x \geq \epsilon^2} \frac{\lambda_x}{\epsilon^2} Q_x \leq \mathrm{Tr} \sum_x \frac{\lambda_x}{\epsilon^2} Q_x \\
&= \frac{1}{\epsilon^2} \mathrm{Tr}\, P_1P_2P_1 = \frac{1}{\epsilon^2} \mathrm{Tr}\, P_1P_2.
\end{aligned}$$

□

Lemma 5 and Lemma 6 below are basic results for subspaces of an inner product space.

Lemma 5. *Let V, W be subspaces of a Hilbert space $\mathcal{H}$, and have direct-sum decompositions $V = \bigoplus_{i=1}^{T_1} V_i$ and $W = \bigoplus_{j=1}^{T_2} W_j$. Suppose we have*

$$\max_{|v\rangle \in V_i, |w\rangle \in W_j} |\langle v|w\rangle| \leq \delta, \qquad \textit{for all } 1 \leq i \leq T_1,\ 1 \leq j \leq T_2.$$

Then the overlap between V and W is bounded as

$$\max_{|v\rangle \in V, |w\rangle \in W} |\langle v|w\rangle| \leq \sqrt{T_1T_2}\delta.$$

Proof. Let $|v\rangle \in V$ and $|w\rangle \in W$ be any two unit vectors. Write $|v\rangle = \sum_i \alpha_i|v_i\rangle$ and $|w\rangle = \sum_j \beta_j|w_j\rangle$, with $|v_i\rangle \in V_i, |w_j\rangle \in W_j$ and $\sum_i |\alpha_i|^2 =$

$\sum_j |\beta_j|^2 = 1$. Making use of the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|\langle v|w\rangle| =& \left|\sum_{i,j} \overline{\alpha_i}\beta_j \langle v_i|w_j\rangle\right| \\
\leq& \sum_{i,j} |\alpha_i| \cdot |\beta_j|\delta \\
=& \Big(\sum_{i=1} |\alpha_i|\Big)\Big(\sum_{j=1} |\beta_j|\Big)\delta \\
\leq& \sqrt{T_1 T_2}\delta,
\end{aligned}$$

and this finishes the proof. □

Lemma 6. *Let $S_1, S_2, \ldots, S_r$ be subspaces of a Hilbert space, such that the overlaps between them are bounded as*

$$\max_{|v_i\rangle \in S_i, |v_j\rangle \in S_j} |\langle v_i|v_j\rangle| \leq \delta, \quad 1 \leq i \neq j \leq r.$$

For all $1 \leq i \leq r$, denote the projector onto S_i as P_i, and denote the projector onto $S = S_1 + S_2 + \cdots + S_r$ as P. Suppose $\delta < \frac{1}{2(r-1)}$. Then

$$P \leq \frac{1-(r-1)\delta}{1-2(r-1)\delta} \sum_{i=1}^{r} P_i. \tag{11}$$

Proof. For an arbitrary unit vector $|v\rangle \in S$, write $|v\rangle = \sum_{i=1}^{r} \alpha_i |v_i\rangle$, with $|v_i\rangle \in S_i$. Then

$$\begin{aligned}
\langle v|\sum_{k=1}^{r} P_k|v\rangle =& \sum_{(i,k):i=k} \sum_j \overline{\alpha_i}\alpha_j \langle v_i|P_k|v_j\rangle + \sum_i \sum_{(j,k):j=k} \overline{\alpha_i}\alpha_j \langle v_i|P_k|v_j\rangle \\
&- \sum_{\substack{(i,j,k):\\ i=j=k}} \overline{\alpha_i}\alpha_j \langle v_i|P_k|v_j\rangle + \sum_k \sum_{i:i\neq k} \sum_{j:j\neq k} \overline{\alpha_i}\alpha_j \langle v_i|P_k|v_j\rangle \\
=& 2\langle v|v\rangle - \sum_{i=1}^{r} |\alpha_i|^2 + \sum_{k=1}^{r} \left(\langle v| - \overline{\alpha_k}\langle v_k|\right) P_k \left(|v\rangle - \alpha_k|v_k\rangle\right).
\end{aligned}$$

The last term is nonnegative. Besides, we have

$$
\begin{aligned}
1 =&\langle v|v\rangle = \sum_{i=1}^{r} |\alpha_i|^2 + \sum_{(i,j):i\neq j} \overline{\alpha_i}\alpha_j \langle v_i|v_j\rangle \\
\geq& \sum_{i=1}^{r} |\alpha_i|^2 - \sum_{(i,j):i\neq j} |\alpha_i| \cdot |\alpha_j| \cdot |\langle v_i|v_j\rangle| \\
\geq& \sum_{i=1}^{r} |\alpha_i|^2 - \sum_{(i,j):i\neq j} \frac{\big(|\alpha_i|^2 + |\alpha_j|^2\big)\,\delta}{2} \\
=&\big(1-(r-1)\delta\big) \sum_{i=1}^{r} |\alpha_i|^2.
\end{aligned}
$$

Combining the above arguments, we get

$$
\langle v| \sum_{k=1}^{r} P_k |v\rangle \geq 2 - \frac{1}{1-(r-1)\delta} = \frac{1-2(r-1)\delta}{1-(r-1)\delta},
$$

which implies equation (11). □

4.2 Construction of the measurements

In the following, we will describe the procedure of constructing a family of projective measurements $\{\Pi_1(\epsilon), \ldots, \Pi_{r-1}(\epsilon), \Pi_r(\epsilon) + \Pi_{r+1}(\epsilon)\}$, which will be used to show that the right-hand side of equation (4) is an achievable error probability.

Our construction is similar to the ones explored in [31, 32] and earlier in [21], especially, in applying the Gram-Schmidt orthonormalization to states that are ordered according to the corresponding eigenvalues, for formulating the projective measurements. However, our method is also significantly different from those of [21, 31, 32]. At first, instead of dealing with every eigenvector of the hypothetic states individually, we treat each of the eigenspaces as a whole. This, for i.i.d. states of the form $\omega^{\otimes n}$, is reminiscent of the method of types [11, 12], from which we indeed benefit. Secondly, we carefully remove from each of these eigenspaces a subspace, in a way such that the perturbation to the hypothetic states is limited but the overlaps between the supporting spaces of them become sufficiently small. As a result, we can effectively employ the Gram-Schmidt process to formulate an approximately optimal measurement.

Recall that for $1 \leq i \leq r$, $A_i = \sum_{k=1}^{T_i} \lambda_{ik} Q_{ik}$ is the spectral decomposition. Let $S_{ik} := \mathrm{supp}(Q_{ik})$ be the eigenspaces of A_i. From now on, we always identify the subscript "ik" with "(i,k)". So λ_{ik}, Q_{ik} and S_{ik} are also denoted as $\lambda_{(i,k)}$, $Q_{(i,k)}$ and $S_{(i,k)}$, respectively. Define the index set $\mathcal{O} := \bigcup_{i=1}^{r} \{(i,k) : k \in \mathbb{N}, 1 \leq k \leq T_i\}$. Now we arrange all the eigenvalues $\{\lambda_{ik}\}_{(i,k)\in\mathcal{O}}$ in a decreasing order, and let $g : \{0, 1, 2, \ldots, |\mathcal{O}|\} \mapsto \{(0,0)\} \cup \mathcal{O}$ be the bijection indicating the position of each λ_{ik} in such an ordering:

$$\lambda_{g(1)} \geq \lambda_{g(2)} \geq \cdots \geq \lambda_{g(|\mathcal{O}|)}, \tag{12}$$

and $g(0) = (0,0)$ is introduced for later convenience. Our construction consists of the following three steps.

Step 1: reducing the overlaps between the eigenspaces. For this purpose, we employ the operation ϵ-subtraction to remove a subspace from each of these eigenspaces. Let $\ominus_\epsilon$ be endowed with a left associativity, that is, $A \ominus_\epsilon B \ominus_\epsilon C := (A \ominus_\epsilon B) \ominus_\epsilon C$. Set $Q_{g(0)} = 0$ and $S_{g(0)} = \{0\}$. We define

$$\tilde{Q}_{g(m)} := \begin{cases} Q_{g(0)}, & \text{if } m = 0 \\ Q_{g(m)} \ominus_\epsilon Q_{g(0)} \ominus_\epsilon Q_{g(1)} \ominus_\epsilon \cdots \ominus_\epsilon Q_{g(m-1)}, & \text{if } 1 \leq m \leq |\mathcal{O}| \end{cases} \tag{13}$$

and

$$\tilde{S}_{g(m)} := \begin{cases} S_{g(0)}, & \text{if } m = 0 \\ S_{g(m)} \ominus_\epsilon S_{g(0)} \ominus_\epsilon S_{g(1)} \ominus_\epsilon \cdots \ominus_\epsilon S_{g(m-1)}, & \text{if } 1 \leq m \leq |\mathcal{O}|. \end{cases} \tag{14}$$

Note that, according to Definition 3, $\tilde{S}_{g(m)} = \mathrm{supp}\left(\tilde{Q}_{g(m)}\right)$. Now we denote

$$\tilde{A}_i := \sum_{k=1}^{T_i} \lambda_{ik} \tilde{Q}_{ik}, \quad 1 \leq i \leq r. \tag{15}$$

We will show later that, for the purpose of the current paper, $\tilde{A}_i$ is a good approximation of A_i.

Step 2: orthogonalizing the eigenspaces. To formulate the projective measurement, we need to assign each of the states $\{A_i\}_i$ an orthogonal subspace, for the projectors to be supported on. To do so, we treat $\tilde{A}_i$'s as representives of A_i's, and orthogonalize the subspaces $\{\tilde{S}_{g(m)}\}_m$, using a Gram-Schmidt process. Define $\hat{S}_{g(0)} := \{0\}$, and for all $1 \leq m \leq |\mathcal{O}|$,

$$\hat{S}_{g(m)} := \left(\tilde{S}_{g(0)} + \cdots + \tilde{S}_{g(m)}\right) \ominus_1 \left(\tilde{S}_{g(0)} + \cdots + \tilde{S}_{g(m-1)}\right), \tag{16}$$

where $\ominus_1$ is the operation "ϵ-subtraction" with $\epsilon = 1$. Recalling Definition 3, we easily see that the right-hand side of equation (16) is just the orthogonal complement of $\tilde{S}_{g(0)} + \cdots + \tilde{S}_{g(m-1)}$, in the space $\tilde{S}_{g(0)} + \cdots + \tilde{S}_{g(m)}$, noticing that obviously the former is a subspace of the latter. So the subspaces $\hat{S}_{g(1)}, \ldots, \hat{S}_{g(m)}$ are mutually orthogonal. Thus the definition in equation (16) is equivalent to

$$\tilde{S}_{g(1)} + \cdots + \tilde{S}_{g(m)} = \bigoplus_{t=1}^{m} \hat{S}_{g(t)}, \quad \text{for all} \;\; 1 \leq m \leq |\mathcal{O}|. \tag{17}$$

Note that it is possible that $\hat{S}_{g(m)} = \{0\}$ for certain values of m, and in these cases we have $\operatorname{proj}\left(\hat{S}_{g(m)}\right) = 0$.

Step 3: defining the family of projective measurements. We set

$$\Pi_i(\epsilon) := \sum_{k=1}^{T_i} \operatorname{proj}\left(\hat{S}_{ik}\right), \quad 1 \leq i \leq r, \tag{18}$$

and let

$$\Pi_{r+1}(\epsilon) := \mathbb{1} - \sum_{i=1}^{r} \Pi_i(\epsilon). \tag{19}$$

Here the parameter ϵ is introduced in step 1. By definition, $\Pi_1(\epsilon), \ldots, \Pi_{r+1}(\epsilon)$ are orthogonal projectors and $\sum_{i=1}^{r+1} \Pi_i(\epsilon) = \mathbb{1}$. So, they form a projective measurement. Our strategy for testing $A_1, \ldots, A_r$ is that, if we get the measurement outcome associated with $\Pi_i(\epsilon)$, we conclude that the state is A_i. For the outcome associated with the extra projector $\Pi_{r+1}(\epsilon)$, we can make any decision, or just report an error; here we simply assign it to A_r. Thus, the family of measurements that we construct for use is

$$\Pi = \left\{\Pi_1(\epsilon), \ldots, \Pi_{r-1}(\epsilon), \Pi_r(\epsilon) + \Pi_{r+1}(\epsilon)\right\}. \tag{20}$$

4.3 Proof of the one-shot achievability bound: Theorem 2

We show that, for properly chosen $\epsilon \in [0, 1]$, the measurement constructed in Section 4.2 will achieve an error probability that equals the right-hand side of equation (4).

Proof of Theorem 2. For the ensemble of nonnegative matrices $\mathcal{A} = \{A_1, \ldots, A_r\}$,

and the measurement Π specified in equation (20), we have

$$\begin{aligned} P_e(\mathcal{A};\Pi) &= \sum_{i=1}^{r-1} \operatorname{Tr} A_i \left(\mathbb{1} - \Pi_i(\epsilon)\right) + \operatorname{Tr} A_r \left(\mathbb{1} - \Pi_r(\epsilon) - \Pi_{r+1}(\epsilon)\right) \\ &\leq \sum_{i=1}^{r} \operatorname{Tr} A_i \left(\mathbb{1} - \Pi_i(\epsilon)\right). \end{aligned}$$

We now make use of the matrices $\{\tilde{A}_i\}$, which are defined in step 1 of the measurement construction in section 4.2; cf. equantion (15). Substituting $(A_i - \tilde{A}_i) + \tilde{A}_i$ for A_i, and noticing that it is an obvious result of equations (13) and (15) that $A_i - \tilde{A}_i \geq 0$, we further bound the error probability as

$$P_e(\mathcal{A};\Pi) \leq \sum_{i=1}^{r} \operatorname{Tr}(A_i - \tilde{A}_i) + \sum_{i=1}^{r} \operatorname{Tr} \tilde{A}_i \left(\mathbb{1} - \Pi_i(\epsilon)\right). \tag{21}$$

In the following, we will evaluate the two terms of the right-hand side of equation (21), separately.

Invoking the spectral decomposition of A_i, and using equation (15), we can write

$$\begin{aligned} \sum_{i=1}^{r} \operatorname{Tr}(A_i - \tilde{A}_i) &= \sum_{i=1}^{r} \sum_{k=1}^{T_i} \lambda_{ik} \operatorname{Tr}(Q_{ik} - \tilde{Q}_{ik}) \\ &= \sum_{m=1}^{|\mathcal{O}|} \lambda_{g(m)} \operatorname{Tr}\left(Q_{g(m)} - \tilde{Q}_{g(m)}\right), \end{aligned} \tag{22}$$

where we have used the map g, introduced in the previous section, to indicate the subscripts. For each integer $2 \leq m \leq |\mathcal{O}|$, applying the third result of Lemma 4 to the ϵ-subtraction formulas

$$\left(Q_{g(m)} \ominus_\epsilon Q_{g(0)} \ominus_\epsilon Q_{g(1)} \ominus_\epsilon \cdots \ominus_\epsilon Q_{g(t-1)}\right) \ominus_\epsilon Q_{g(t)}, \quad 1 \leq t \leq m-1,$$

we obtain

$$\begin{aligned} &\operatorname{Tr}\left(Q_{g(m)} \ominus_\epsilon Q_{g(0)} \ominus_\epsilon \cdots \ominus_\epsilon Q_{g(t-1)}\right) - \operatorname{Tr}\left(Q_{g(m)} \ominus_\epsilon Q_{g(0)} \ominus_\epsilon \cdots \ominus_\epsilon Q_{g(t)}\right) \\ &\leq \frac{1}{\epsilon^2} \operatorname{Tr}\left(Q_{g(m)} \ominus_\epsilon Q_{g(0)} \ominus_\epsilon \cdots \ominus_\epsilon Q_{g(t-1)}\right) Q_{g(t)} \\ &\leq \frac{1}{\epsilon^2} \operatorname{Tr} Q_{g(m)} Q_{g(t)}, \end{aligned} \tag{23}$$

where for the last inequality we have used repeatedly the first result of Lemma 4. Summing over $t \in \{1, 2, \ldots, m-1\}$ at both the first and the last line of Equation (23), yields

$$\begin{aligned} &\operatorname{Tr} Q_{g(m)} - \operatorname{Tr}\left(Q_{g(m)} \ominus_\epsilon Q_{g(0)} \ominus_\epsilon Q_{g(1)} \ominus_\epsilon \cdots \ominus_\epsilon Q_{g(m-1)}\right) \\ \leq &\frac{1}{\epsilon^2} \sum_{t=1}^{m-1} \operatorname{Tr} Q_{g(m)} Q_{g(t)}, \end{aligned} \tag{24}$$

for all $2 \leq m \leq |\mathcal{O}|$. Combining equations (13) and (24), and directly verifying the case $m = 1$, we get

$$\operatorname{Tr}\left(Q_{g(m)} - \tilde{Q}_{g(m)}\right) \leq \frac{1}{\epsilon^2} \sum_{t=0}^{m-1} \operatorname{Tr} Q_{g(m)} Q_{g(t)}, \quad 1 \leq m \leq |\mathcal{O}|. \tag{25}$$

Eventually, inserting equation (25) into equation (22), and changing the subscripts, we arrive at

$$\begin{aligned} \sum_{i=1}^{r} \operatorname{Tr}(A_i - \tilde{A}_i) \leq &\frac{1}{\epsilon^2} \sum_{m=1}^{|\mathcal{O}|} \sum_{t=0}^{m-1} \lambda_{g(m)} \operatorname{Tr} Q_{g(m)} Q_{g(t)} \\ = &\frac{1}{\epsilon^2} \sum_{(i,j):i<j} \sum_{k,\ell} \min\{\lambda_{ik}, \lambda_{j\ell}\} \operatorname{Tr} Q_{ik} Q_{j\ell}. \end{aligned} \tag{26}$$

Now we evaluate the second term of the right-hand side of equation (21). Using equations (15) and (18), and employing the map g to indicate the subscripts, we can write

$$\begin{aligned} \sum_{i=1}^{r} \operatorname{Tr} \tilde{A}_i \left(\mathbb{1} - \Pi_i(\epsilon)\right) = &\sum_{i=1}^{r} \sum_{k=1}^{T_i} \lambda_{ik} \operatorname{Tr} \tilde{Q}_{ik} \left(\mathbb{1} - \sum_{\ell=1}^{T_i} \operatorname{proj}\left(\hat{S}_{i\ell}\right)\right) \\ \leq &\sum_{i=1}^{r} \sum_{k=1}^{T_i} \lambda_{ik} \operatorname{Tr} \tilde{Q}_{ik} \left(\mathbb{1} - \operatorname{proj}\left(\hat{S}_{ik}\right)\right) \\ = &\sum_{m=1}^{|\mathcal{O}|} \lambda_{g(m)} \operatorname{Tr} \tilde{Q}_{g(m)} \left(\mathbb{1} - \operatorname{proj}\left(\hat{S}_{g(m)}\right)\right). \end{aligned} \tag{27}$$

Equation (17) implies that

$$\operatorname{supp}\left(\tilde{Q}_{g(m)}\right) = \tilde{S}_{g(m)} \subseteq \bigoplus_{t=1}^{m} \hat{S}_{g(t)}.$$

As a result, the identity matrix in the third line of equation (27) can be replaced by $\sum_{t=1}^{m} \mathrm{proj}\left(\hat{S}_{g(t)}\right)$. This gives

$$\begin{aligned}\sum_{i=1}^{r} \mathrm{Tr}\,\tilde{A}_i\left(\mathbb{1}-\Pi_i(\epsilon)\right) &\leq \sum_{m=1}^{|\mathcal{O}|} \lambda_{g(m)}\,\mathrm{Tr}\,\tilde{Q}_{g(m)} \sum_{t=0}^{m-1} \mathrm{proj}\left(\hat{S}_{g(t)}\right) \\ &= \sum_{m=1}^{|\mathcal{O}|} \lambda_{g(m)}\,\mathrm{Tr}\,\tilde{Q}_{g(m)}\,\mathrm{proj}\left(\sum_{t=0}^{m-1} \tilde{S}_{g(t)}\right),\end{aligned} \tag{28}$$

where we have $\hat{S}_{g(0)} = \tilde{S}_{g(0)} = \{0\}$, and for the equality we used equation (17). The next step is to upper bound $\mathrm{proj}\left(\sum_{t=0}^{m-1} \tilde{S}_{g(t)}\right)$, with a quantity in terms of $\sum_{t=0}^{m-1} \tilde{Q}_{g(t)}$. This can be done by directly applying Lemma 6. However, we notice that, for each $1 \leq i \leq r$ the subspaces in the set $\{\tilde{S}_{ik}\}_k$ are orthogonal and we can make use of this fact to derive a tighter bound. For a pair of numbers (x, y), let $[(x, y)]_1$ denote the first component: $[(x, y)]_1 = x$. We write

$$\sum_{t=0}^{m-1} \tilde{S}_{g(t)} = \sum_{i=1}^{r} \tilde{S}_i^m, \quad \text{with } \tilde{S}_i^m := \bigoplus_{\substack{t:0\leq t\leq m-1,\\ [g(t)]_1=i}} \tilde{S}_{g(t)}. \tag{29}$$

We will use Lemma 5 to bound the overlaps between each pair of the subspaces $\{\tilde{S}_i^m\}_{i=1}^r$, and then we apply Lemma 6. Although we will only get a slightly better bound, compared to applying Lemma 6 directly, it is possible to make this improvement bigger by strengthening the result of Lemma 6. Now, due to Lemma 4 and the definition of $\tilde{S}_{g(t)}$ [cf. equation (14)], we can bound

$$\max_{|v\rangle\in\tilde{S}_{g(t)},|v'\rangle\in S_{g(t')}} \left|\langle v|v'\rangle\right| \leq \epsilon, \quad 0 \leq t' < t \leq |\mathcal{O}|.$$

Since $\tilde{S}_{g(t')}$ is a subspace of $S_{g(t')}$, it follows that

$$\max_{|v\rangle\in\tilde{S}_{g(t)},|v'\rangle\in\tilde{S}_{g(t')}} \left|\langle v|v'\rangle\right| \leq \epsilon, \quad 0 \leq t \neq t' \leq |\mathcal{O}|.$$

Recalling the spectral decomposition of each A_i, we notice that the direct sum in equation (29) has at most T_i terms. So, an application of Lemma 5 gives us that, for every $1 \leq m \leq |\mathcal{O}|$,

$$\max_{|v\rangle\in\tilde{S}_i^m,|w\rangle\in\tilde{S}_j^m} |\langle v|w\rangle| \leq T\epsilon, \quad 1 \leq i \neq j \leq r, \tag{30}$$

where $T = \max\{T_1, \ldots, T_r\}$. Equation (30) allows us to apply Lemma 6 and obtain

$$\operatorname{proj}\left(\sum_{t=0}^{m-1} \tilde{S}_{g(t)}\right) \le \frac{1-(r-1)T\epsilon}{1-2(r-1)T\epsilon} \sum_{t=0}^{m-1} \tilde{Q}_{g(t)}\,, \quad 1 \le m \le |\mathcal{O}|, \tag{31}$$

for which we have also used equation (29) and the fact that $\operatorname{proj}\left(\tilde{S}_{g(t)}\right) = \tilde{Q}_{g(t)}$. Note that here the condition on δ in Lemma 6 is satisfied for our later choice of ϵ. Now, inserting equation (31) into equation (28), and making use of the relation $\tilde{Q}_{g(m)} \le Q_{g(m)}$ for all $0 \le m \le |\mathcal{O}|$, we arrive at

$$\sum_{i=1}^{r} \operatorname{Tr} \tilde{A}_i \left(\mathbb{1} - \Pi_i(\epsilon)\right) \le \frac{1-(r-1)T\epsilon}{1-2(r-1)T\epsilon} \sum_{m=1}^{|\mathcal{O}|} \lambda_{g(m)} \sum_{t=0}^{m-1} \operatorname{Tr} Q_{g(m)} Q_{g(t)}, \tag{32}$$

which translates to

$$\sum_{i=1}^{r} \operatorname{Tr} \tilde{A}_i \left(\mathbb{1} - \Pi_i(\epsilon)\right) \le \frac{1-(r-1)T\epsilon}{1-2(r-1)T\epsilon} \sum_{(i,j):i<j} \sum_{k,\ell} \min\{\lambda_{ik}, \lambda_{j\ell}\} \operatorname{Tr} Q_{ik} Q_{j\ell}. \tag{33}$$

Eventually, inserting equations (26) and (33) into equation (21)and setting $\epsilon = \frac{2}{5(r-1)T}$ lets us obtain equation (4), with

$$f(r,T) = \frac{25(r-1)^2 T^2}{4} + 3 < 10(r-1)^2 T^2,$$

and we are done. □

4.4 Proof of the error exponent

We are now ready for the proof of Theorem 1.

Proof of Theorem 1. For the achievability part, we use Theorem 2. Let $d = |\mathcal{H}|$ be the dimension of the associated Hilbert space of the states $\rho_1, \ldots, \rho_r$. The type counting lemma (see, e.g. [10], Theorem 12.1.1) provides that the number of eigenspaces of the states $\rho_1^{\otimes n}, \ldots, \rho_r^{\otimes n}$ satisfies

$$\Omega\left(\rho_i^{\otimes n}\right) \le (n+1)^d, \quad \forall\, 1 \le i \le r.$$

For all $1 \le i \le r$, let $\rho_i^{\otimes n} = \sum_k \lambda_{ik}^{(n)} Q_{ik}^{(n)}$ be written in the spectral decomposition. Theorem 2 gives

$$\begin{aligned} &P_e^*\left(\{p_1\rho_1^{\otimes n}, \ldots, p_r\rho_r^{\otimes n}\}\right) \\ &\le 10(r-1)^2(n+1)^{2d} \sum_{(i,j):i<j} \sum_{k,\ell} \min\{p_i\lambda_{ik}^{(n)}, p_j\lambda_{j\ell}^{(n)}\} \operatorname{Tr} Q_{ik}^{(n)} Q_{j\ell}^{(n)}. \end{aligned} \tag{34}$$

Furthermore, for any $1 \leq i < j \leq r$, we have

$$
\begin{aligned}
&\sum_{k,\ell} \min\{p_i\lambda_{ik}^{(n)}, p_j\lambda_{j\ell}^{(n)}\} \operatorname{Tr} Q_{ik}^{(n)} Q_{j\ell}^{(n)} \\
\leq &\max\{p_i, p_j\} \min_{0\leq s\leq 1} \sum_{k,\ell} \big(\lambda_{ik}^{(n)}\big)^s \big(\lambda_{j\ell}^{(n)}\big)^{1-s} \operatorname{Tr} Q_{ik}^{(n)} Q_{j\ell}^{(n)} \\
= &\max\{p_i, p_j\} \min_{0\leq s\leq 1} \left(\operatorname{Tr} \rho_i^s \rho_j^{1-s}\right)^n .
\end{aligned}
\tag{35}
$$

Inserting equation (35) into equation (34), together with some basic calculus, results in

$$
\begin{aligned}
&P_e^*\left(\{p_1\rho_1^{\otimes n}, \ldots, p_r\rho_r^{\otimes n}\}\right) \\
\leq &10(r-1)^2 C_r^2 (n+1)^{2d} \max\{p_1, \ldots, p_r\} \max_{(i,j):i\neq j} \min_{0\leq s\leq 1} \left(\operatorname{Tr} \rho_i^s \rho_j^{1-s}\right)^n ,
\end{aligned}
\tag{36}
$$

Where $C_r^2 = \frac{r(r-1)}{2}$ is a binomial coefficient. From equation (36) we easily derive

$$
\liminf_{n\to\infty} \frac{-1}{n} \log P_e^*\left(\{p_1\rho_1^{\otimes n}, \ldots, p_r\rho_r^{\otimes n}\}\right) \geq \min_{(i,j):i\neq j} \max_{0\leq s\leq 1} \left\{-\log \operatorname{Tr} \rho_i^s \rho_j^{1-s}\right\} .
\tag{37}
$$

On the other hand, the optimality part, that

$$
\limsup_{n\to\infty} \frac{-1}{n} \log P_e^*\left(\{p_1\rho_1^{\otimes n}, \ldots, p_r\rho_r^{\otimes n}\}\right) \leq \min_{(i,j):i\neq j} \max_{0\leq s\leq 1} \left\{-\log \operatorname{Tr} \rho_i^s \rho_j^{1-s}\right\} ,
\tag{38}
$$

is a straightforward generalization of the $r = 2$ situation [29]; see [31] for the proof. Alternatively, one can start with the one-shot bound of equation (5). Then we use the fact that equation (7), when applied to the i.i.d. states and acted by "$\frac{-1}{n}\log$" at both sides, becomes asymptotically an equality. Note that this is still based on the results of Nussbaum and Szkoła in [29].

At last, equation (37) and equation (38) together are obviously equivalent to equation (3) and we conclude the proof of Theorem 1. □

5 Discussion

By explicitly constructing a family of asymptotically optimal measurements for testing quantum hypotheses $\{\rho_1^{\otimes n}, \ldots, \rho_r^{\otimes n}\}$, we have proven the achievability of the multiple quantum Chernoff distance, and eventually established that this is the optimal rate exponent at which the error decays.

In the nonasymptotic setting, we have obtained a new upper bound for the optimal average error probability in discriminating a set of density matrices $\{A_1, \ldots, A_r\}$, which satisfy $A_i \geq 0$ and are not necessarily normalized. Yuen, Kennedy and Lax [41] derived a formula for the optimal average error:

$$P_e^*(A_1, \ldots, A_r) = \mathrm{Tr} \sum_i A_i - \min \{\mathrm{Tr}\, X : X \geq A_i,\ i = 1, \ldots, r\}\,; \qquad (39)$$

see also [22] and [2] for alternative formulations. However, the fact that equation (39) involves an optimization problem itself, makes it difficult to apply this formula directly. Our upper bound stated in Theorem 2, though looser compared to equation (39), has an explicit form and there is a dual lower bound as shown in equation (5). We thus hope that it will find more applications.

We wonder whether the states-dependent factor $f(r, T)$ can be replaced by a constant, or at least can be improved such that it only depends on r (see also a similar conjecture made in [2]). While it is possible that we can improve Lemma 6 to give a better bound on $f(r, T)$, we do not think that this can remove the dependence on T and r. In this direction, the pretty good measurement [4, 15] and its variant [40], both of which achieve an error probability lying between P_e^* and $2P_e^*$, may be useful tools to try. In fact, in Theorem 2 the dependence of our bound on T is not necessary: using the argument in [38], we can convert it into a dependence on the relation between the maximal and the minimal eigenvalues of the hypothetic states; see Proposition 7 below and the proof in the Appendix. This conversion is useful when the spectrum of each A_i is sufficiently flat, no matter how big the number of their eigenspaces is.

Proposition 7. *For all $i = 1, \ldots, r$, let $\lambda_{max}(A_i)$ be the maximal eigenvalue, and $\lambda_{min}(A_i)$ be the minimal nonzero eigenvalue of A_i. Denote*

$$L := \max \left\{ \left\lfloor \log_2 \frac{2\lambda_{max}(A_1)}{\lambda_{min}(A_1)} \right\rfloor, \ldots, \left\lfloor \log_2 \frac{2\lambda_{max}(A_r)}{\lambda_{min}(A_r)} \right\rfloor \right\}.$$

Then, in Theorem 2 the states-dependent factor $f(r, T)$ can be replaced by $h(r, L) := 40(r-1)^2 L^2$.

Another interesting question is how our method can be extended to deal with the problem of discriminating correlated states, where each of the hypothetic states $\rho_1^{(n)}, \ldots, \rho_r^{(n)}$ can be correlated among the n subsystems. The upper bound stated in Theorem 2 (also in Proposition 7 for an alternative states-dependent factor), together with the dual lower bound of equation (5),

can be used to analyse the asymptotic behavior of the error. This method may identify the optimal error exponent which can be quite different from reasonable generalizations of the Chernoff distance, in contrast to previous works which under certain conditions yield the mean quantum Chernoff distance; see, for example, [17, 25, 26, 30]. However, the main difficulty we will confront in this method is to characterize the spectral decomposition of the correlated states when n goes to infinity. At last, a particularly interesting problem in this setting, proposed by Audenaert and Mosonyi [2], is testing composite hypotheses, say, $\rho^{\otimes n}$ versus $\sum_i q_i \sigma_i^{\otimes n}$. Here the sum may be replaced by an integral. See also discussions in [5] and [7] of this problem in the asymmetric case of Stein's lemma. While our method for proving Theorem 1 does shed some light on this problem, it seems that a complete solution needs further ideas.

Acknowledgments. The author is grateful to Charles Bennett, Aram Harrow, Graeme Smith, John Smolin and Andreas Winter, for very helpful discussions and comments. He further thanks Fernando Brandão, Milán Mosonyi, Marco Tomamichel and the anonymous referees for their enthusiasm and/or helpful comments.

Appendix A: proof of Proposition 7

Proof. For an arbitrary nonnegative matrix $A = \sum_k \lambda_k Q_k$ written in the spectral decomposition form, define the modified version of A as

$$A' = \sum_{m=1}^{M} 2^m \lambda_{min}(A) \sum_{k:\lambda_k \in \mathcal{O}_m} Q_k,$$

where $M := \left\lfloor \log_2 \frac{2\lambda_{max}(A)}{\lambda_{min}(A)} \right\rfloor = \Omega(A')$ and $\mathcal{O}_m := \{\lambda_k : 2^{m-1}\lambda_{min}(A) \leq \lambda_k < 2^m \lambda_{min}(A)\}$. Then we have $A \leq A' \leq 2A$, and also A and A' commute. Now for $A_1, \ldots, A_r$, we define $A'_1, \ldots, A'_r$ in a similarly way as A' was defined. Obviously, $\Omega(A'_i) = \left\lfloor \log_2 \frac{2\lambda_{max}(A_i)}{\lambda_{min}(A_i)} \right\rfloor$. Applying Theorem 2, we can evaluate

$$\begin{aligned} & P_e^* \left(\{A'_1, \ldots, A'_r\}\right) \\ \leq & 10(r-1)^2 L^2 \cdot 4 \sum_{(i,j):i<j} \sum_{k,\ell} \min\{\lambda_{ik}, \lambda_{j\ell}\} \operatorname{Tr} Q_{ik} Q_{j\ell}. \end{aligned} \tag{40}$$

On the other hand, since for all i, $A_i \leq A'_i$, we have by the definition of P_e^* that

$$P_e^* \left(\{A_1, \ldots, A_r\}\right) \leq P_e^* \left(\{A'_1, \ldots, A'_r\}\right). \tag{41}$$

Equations (40) and (41) together lead to the advertised result. $\square$

References

[1] Audenaert, K. M. R., Casamiglia, J., Munoz-Tapia, R., Bagan, E., Masanes, Ll., Acin, A. and Verstraete, F. (2007). Discriminating states: the quantum Chernoff bound. *Phys. Rev. Lett.* **98** 160501. arXiv:quant-ph/0610027.

[2] Audenaert, K. M. R. and Mosonyi, M. (2014). Upper bounds on the error probabilities and asymptotic error exponents in quantum multiple state discrimination. *J. Math. Phys.* **55** 102201.

[3] Audenaert, K. M. R., Nussbaum, M., Szkoła, A. and Verstraete, F. (2008). Asymptotic error rates in quantum hypothesis testing. *Comm. Math. Phys.* **279** (1) 251–283.

[4] Barnum, H. and Knill, E. (2002). Reversing quantum dynamics with near-optimal quantum and classical fidelity. *J. Math. Phys.* **43** 2097–2106.

[5] Bjelaković, I., Deuschel, J. D., Krüger, T., Seiler, R., Siegmund-Schultze, Ra. and Szkoła, A. (2005). A quantum version of Sanov's theorem. *Comm. Math. Phys.* **260** (3) 659–671.

[6] Blahut, R. E. (1974). Hypothesis testing and information theory. *IEEE Trans. Inf. Theory* **20** (4) 405–417.

[7] Brandão, F. G. S. L., Harrow, A. W., Oppenheim, J. and Strelchuk, S. (2015). Quantum conditional mutual information, reconstructed states, and state redistribution. *Phys. Rev. Lett.* **115** 050501.

[8] Brandão, F. G. S. L. and Plenio, M. B. (2010). A generalization of quantum Stein's lemma. *Comm. Math. Phys.* **295** (3) 791–828.

[9] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** (4) 493–507.

[10] Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley Series in Telecommunications, John Wiley & Sons, New York.

[11] Csiszár, I. (1998). The method of types. *IEEE Trans. Inf. Theory* **44** (6) 2505–2523.

[12] Csiszár, I. and Körner, J. (1981). *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York.

[13] Csiszár, I. and Longo, G. (1971). On the error exponent for source coding and for testing simple statistical hypotheses. *Studia Sci. Math. Hungarica* **6** 181–191.

[14] Han, T. S. and Kobayashi, K. (1989). The strong converse theorem for hypothesis testing. *IEEE Trans. Inf. Theory* **35** (1) 178–180.

[15] Hausladen, P. and Wootters, W. (1994). A "pretty good" measurement for disdinguishing quantum states. *J. Mod. Opt.* **41** 2385.

[16] Helstrom, C. W. (1976). *Quantum Detection and Estimation Theory.* Academic Press, New York.

[17] Hiai, F., Mosonyi, M. and Ogawa, T. (2007). Large deviations and Chernoff bound for certain correlated states on a spin chain. *J. Math. Phys.* **48** 123301.

[18] Hiai, F. and Petz, D. (1991). The proper formula for relative entropy and its asymptotics in quantum probability. *Comm. Math. Phys.* **143** (1) 99–114.

[19] Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** (2) 369–401.

[20] Holevo, A. S. (1973). Statistical decision theory for quantum systems. *J. Multivariate Anal.* **3** 337–394.

[21] Holevo, A. S. (1978). On asymptotically optimal hypothesis testing in quantum statistics. *Theor. Prob. Appl.* **23** 411–415.

[22] König, R., Renner, R. and Schaffner, C. (2009). The operational meaning of min- and max-entropy. *IEEE Trans. Inf. Theory* **55** (9) 4337–4347.

[23] Leang, C. C. and Johnson, D. H. (1997). On the asymptotics of M-hypothesis Bayesian detection. *IEEE Trans. Inf. Theory* **43** (1) 280–282.

[24] Li, K. (2014). Second-order asymptotics for quantum hypothesis testing. *Ann. Statist.* **42** (1) 171–189.

[25] Mosonyi, M. (2009). Hypothesis testing for Gaussian states on bosonic lattices. *J. Math. Phys.* **50** 032105.

[26] Mosonyi, M., Hiai, F., Ogawa, T. and Fannes, M. (2008). Asymptotic distinguishability measures for shift-invariant quasi-free states of fermionic lattice systems. *J. Math. Phys.* **49** 072104.

[27] Mosonyi, M. and Ogawa, T. (2015). Quantum hypothesis testing and the operational interpretation of the quantum Renyi relative entropies. *Comm. Math. Phys.* **334** (3) 1617–1648.

[28] Nussbaum, M. (2013). Attainment of the multiple quantum Chernoff bound for certain ensembles of mixed states. In *Proceedings of the First International Workshop on Entangled Coherent States and Its Application to Quantum Information Science, Tamagawa University, Tokyo, Japan* 77–81.

[29] Nussbaum, M. and Szkoła, A. (2009). The Chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Statist.* **37** (2) 1040–1057. arXiv:quant-ph/0607216.

[30] Nussbaum, M. and Szkoła, A. (2010). Exponential error rates in multiple state discrimination on a quantum spin chain. *J. Math. Phys.* **51** 072203.

[31] Nussbaum, M. and Szkoła, A. (2011). Asymptotically optimal discrimination between multiple pure quantum states. In *Theory of Quantum Computation, Communication and Cryptography. 5th Conference, TQC 2010, Leeds, UK. Lecture Notes in Computer Science* **6519** 1–8. Springer, Berlin.

[32] Nussbaum, M. and Szkoła, A. (2011). An asymptotic error bound for testing multiple quantum hypotheses. *Ann. Statist.* **39** (6) 3211–3233.

[33] Ogawa, T. and Nagaoka, H. (2000). Strong converse and Stein's lemma in the quantum hypothesis testing. *IEEE Trans. Inf. Theory* **46** (7) 2428–2433.

[34] Parthasarathy, K. R. (2001). On consistency of the maximum likelihood method in testing multiple quantum hypotheses. In *Stochastics in Finite and Infinite Dimensions* 361–377. Birkhäuser, Boston.

[35] Qiu, D. W. (2008). Minimum-error discrimination between mixed quantum states. *Phys. Rev. A* **77** 012328.

[36] SALIHOV, N. P. (1973). Asymptotic properties of error probabilities of tests for distinguishing between several multinomial testing schemes. *Dokl. Akad. Nauk SSSR* **209** 54–57.

[37] SALIHOV, N. P. (1998). On a generalization of Chernoff distance. *Teor. Veroyatn. Primen.* **43** 294–314. Translation in *Theory Probab. Appl.* **43** (1999) 239–255.

[38] TOMAMICHEL, M. and HAYASHI, M. (2013). A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Trans. Inf. Theory* **59** (11) 7693–7710.

[39] TORGERSEN, E. N. (1981). Measures of information based on comparison with total information and with total ignorance. *Ann. Statist.* **9** 638–657.

[40] TYSON, J. (2009). Two-sided estimates of minimum-error distinguishability of mixed quantum states via generalized Holevo-Curlander bounds. *J. Math. Phys.* **50** 032106.

[41] YUEN, H. P., KENNEDY, R. S. and LAX, M. (1975). Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Trans. Inf. Theory* **21** (2) 125–134.